Volume 13 Number 6 1985 Nucleic Acids Research

Nucleotide sequence of a functional cDNA for human thymidylate synthase

Keiichi Takeishi, Sumiko Kaneda, Dai Ayusawa, Kimiko Shimizu, Osamu Gotoh* and Takeshi Seno

Department of Immunology and Virology, and *Department of Biochemistry, Saitama Cancer Center Research Institute, Ina-machi, Saitama-ken 362, Japan

Received 7 February 1985; Accepted 1 March 1985

ABSTRACT

We have determined the nucleotide sequence of a cDNA clone, pcHTS-1, encoding human thymidylate synthase (5,10-methylenetetrahydrofolate: dUMP C-methyltransferase, EC 2.1.1.45) which was previously isolated from a human fibroblast expressible cDNA library and functional in mouse cells. The 1.6 kilobase cDNA insert of pcHTS-1 encodes a subunit protein of 313 amino acid (Mr=35,706) and its predicted amino acid sequence is highly conserved in many regions including folylpolyglutamate and 5-fluoro-2'-deoxyuridylate binding sites, when compared with those of Lactobacillus casei, Escherichia coli, and bacteriophage T4. The cDNA contains in its 5'-untranslated region a triple tandemly repeated sequence consisting of 90 nucleotides, which starts immediately upstream of the ATG initiator codon, is very high in G+C content (80%), and can form three possible interconvertible stem-loop structures.

INTRODUCTION

Thymidylate synthase (5,10-methylenetetrahydrofolate: dUMP C-methyltransferase, EC 2.1.1.45) catalyzes the conversion of deoxyuridylate to thymidylate. The enzyme in eukaryotic cells is of particular interest not only for its pivotal role in DNA biosynthesis in relation to the cell cycle (1,2), but also for its involvement in induction of DNA double-strand breaks (3) and heritable fragile sites associated with mental retardation (4) and possibly with certain chromosomal rearrangements in neoplastic cells (5-7). To determine the structure of the thymidylate synthase gene and the regulation of its expression at a molecular level, we recently cloned genomic DNA segments partially encoding human thymidylate synthase from a mouse cell transformant into λ phage vector (8,9), and more recently isolated seven cDNA clones specifying human thymidylate synthase (10) from a human fibroblast cDNA library constructed by Okayama and Berg (11). Two of them were shown to have activity to stably transform mouse thymidylate synthase-deficient cells by expressing human thymidylate synthase (10). The present report describes the nucleotide sequence of the 1.6 kilobase cDNA in one of the clones, pcHTS-1, which encodes

a subunit protein of 313 amino acids and contains a unique tandemly repeated sequence in the 5'-untranslated region. The predicted amino acid sequence as the first eukaryotic case was compared with those of prokaryotes.

MATERIALS AND METHODS

Enzymes, Chemicals and Strains

Restriction endonucleases were purchased from Takara Shuzo (Kyoto, Japan) and New England BioLabs. T4 DNA ligase was obtained from Boehringer-Mannheim. [α-^{32}P]dCTP (400 Ci/mmol, 1 Ci=37GBq), M13 Cloning Kits and M13 Sequencing Kits were from Amersham. A M13 Cloning Kit including M13 RF DNA mp10 and M13 RF DNA mp11 as cloning vectors and Escherichia coli K12 strain JM105 (Δlac-pro, thi, strA, endA, sbcB15, hsdR4, F'traD36, proAB, lacIZΔm15) as a host strain was used in this study.

DNA Isolation

Plasmid DNA of a cDNA clone pHTS-1 (10) was prepared from chloramphenicol-amplified cultures by detergent lysis (12). DNA restriction endonuclease fragments were purified by agarose gel electrophoresis and recovered from the agarose by the glass powder method (13).

DNA Sequence Analysis

A 1.9 kilobase XhoI fragment containing the cDNA insert was isolated in high purity from the XhoI digest of pcHTS-1 DNA. The XhoI fragment was digested with restriction endonucleases, Sau3A, AluI, and PstI, and the resulting fragments were subcloned into an appropriate site of M13 vectors using M13 Cloning Kits according to the method of Messing (14) as described in the "M13 cloning and sequencing handbook(Amersham)". The cloned recombinant M13 phage DNAs were sequenced by the dideoxy chain termination method (15) using thin acrylamide gels (16), as described in the same handbook. Homologous and complementary regions in DNA sequences were examined using the computor program of Queen and Korn (17).

RESULTS AND DISCUSSION

Nucleotide Sequence

A human thymidylate synthase cDNA clone pcHTS-1 which was functional in the thymidylate synthase-negative mouse cells (10) was subjected to sequence analysis. The restriction map of pcHTS-1 indicated that a 1.9 kilobase fragment excised from the plasmid by XhoI digestion contained an entire cDNA insert corresponding to human thymidylate synthase mRNA (10). The nucleotide sequence of the XhoI fragment was determined by the dideoxynucleotide chain

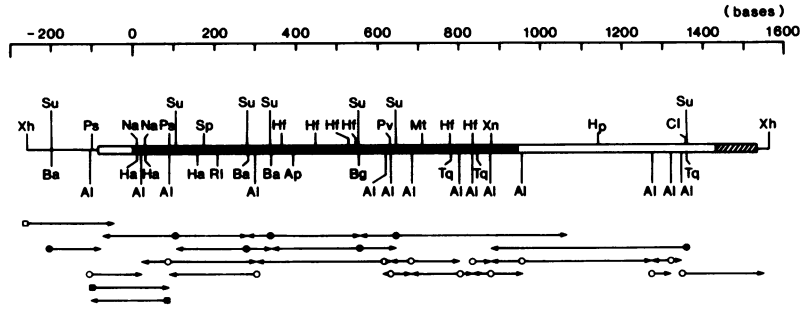

Fig. 1. Restriction map of the cDNA insert in a human thymidylate synthase cDNA clone pcHTS-1 and location of fragments used to determine the nucleotide sequence. The top scale indicates the positions of nucleotides (in bases) relative to the protein initiator codon beginning at position 1. The restriction map shown in the second line was obtained from the sequenced structure of the cDNA insert. The protein coding region and untranslated regions are indicated as a solid bar and open bars, respectively. Poly(A) is shown as a hatched region. The flanking vector sequence regions are indicated by thin lines. The arrows indicate the direction and extent of sequence determination for each fragment analyzed. Symbols at the ends of the arrows indicate the following fragments that were subcloned in M13 vectors and sequenced by the dideoxy chain termination method (15): □ , a XhoI fragment containing the cDNA insert which was subcloned in the SalI site of M13 mp11; ●, fragments produced by Sau3A digestion of the XhoI fragment and subcloned in the BamHI site of M13 mp10; ○, fragments produced by AluI digestion of the XhoI fragment and subcloned in the SmaI site of M13 mp10; ■, fragment produced by PstI digestion of the XhoI fragment and subcloned in the PstI site of M13 mp10. Restriction sites: AT, AluI; Ap, ApaI; Ba, BamHI; Bg, BglII; Cl, ClaI; Ha, HpaII; Hf, HinfI; Hp, HpaI; Mt, MstI; Na, NaeI, Ps, PstI; Pv, PvuII; RI, EcoRI; Sp, SphI; Su, Sau3A; Tq, TagI; Xh, XhoI; Xn, XmnI.

termination method (15). The fragments produced by digestion with Sau3A, AluI, and PstI were subcloned with M13 cloning vectors (14), and the resulting cloned DNAs were sequenced by the strategy illustrated in Fig. 1. Almost all regions of both DNA strands were sequenced. The sequence thus determined is shown in Fig. 2. The cDNA insert consisted of 1,524 nucleotides with 12 deoxyguanylate residues in an oligo(dG) tail at the 5'-end and about 100 deoxyadenylate residues in a poly(A) tract at the 3'-end. It is not known whether an A residue adjacent to the oligo(dG) tail is a transcription initiation site. The longest open reading frame begins with the methionine codon at nucleotide positions 1-3 and extends over a stretch of 939 base pairs. This reading frame can encode a protein of 313 amino acids with a Mr of 35,706, which is close to the Mr (about 33,000) of human thymidyalte synthase determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (18, 19). The sequence from amino acid residue 2 to 25 deduced from this reading

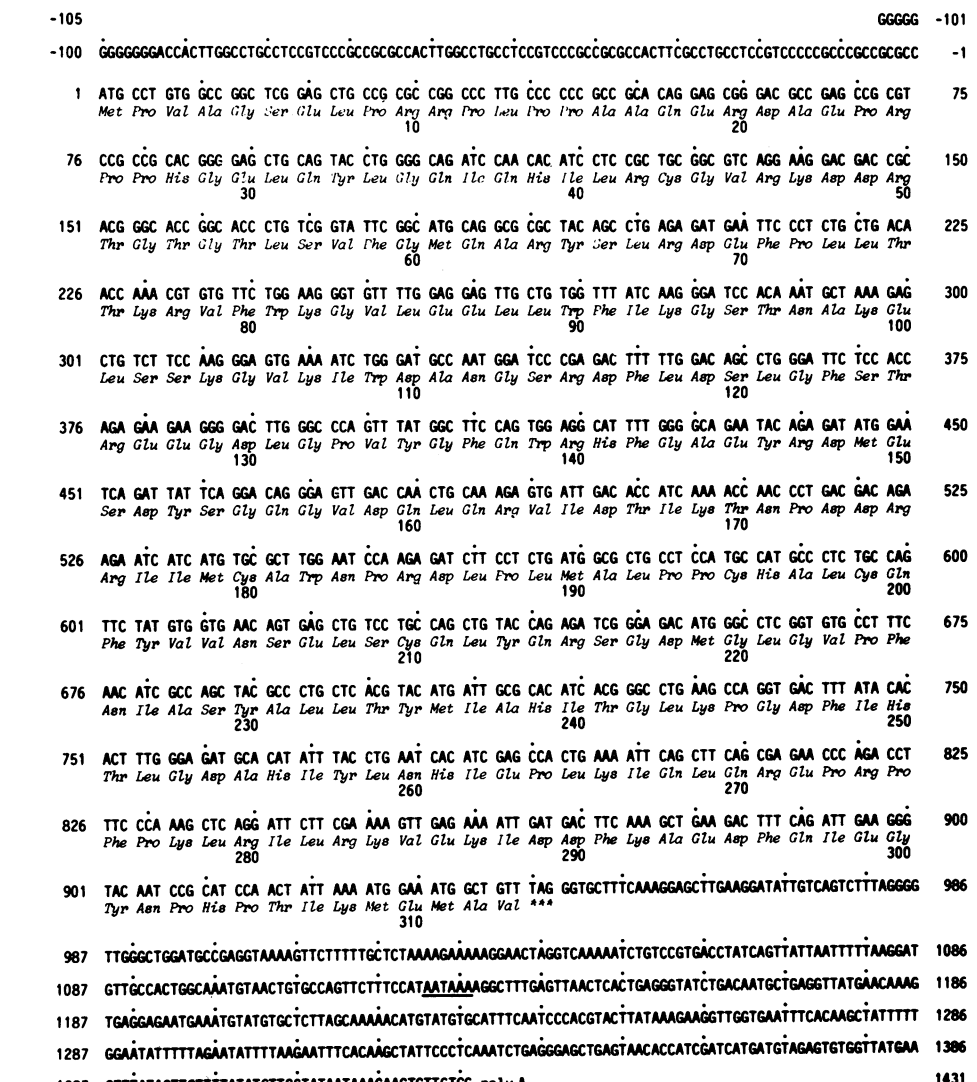

Fig. 2. Complete nucleotide sequence of the cDNA insert in pcHTS-1. Nucleotides are numbered in the 5' to 3' direction. Position 1 corresponds to the first nucleotide of the ATG triplet coding for the initiator methionine. The nucleotides on the 5' side of position 1 are indicated by negative numbers. The predicted amino acid sequence of human thymidylate synthase is shown in italics with asterisks at the stop codon. Polyadenylation signals are underlined. The amino acid residues are numbered beginning with the initiator methionine.

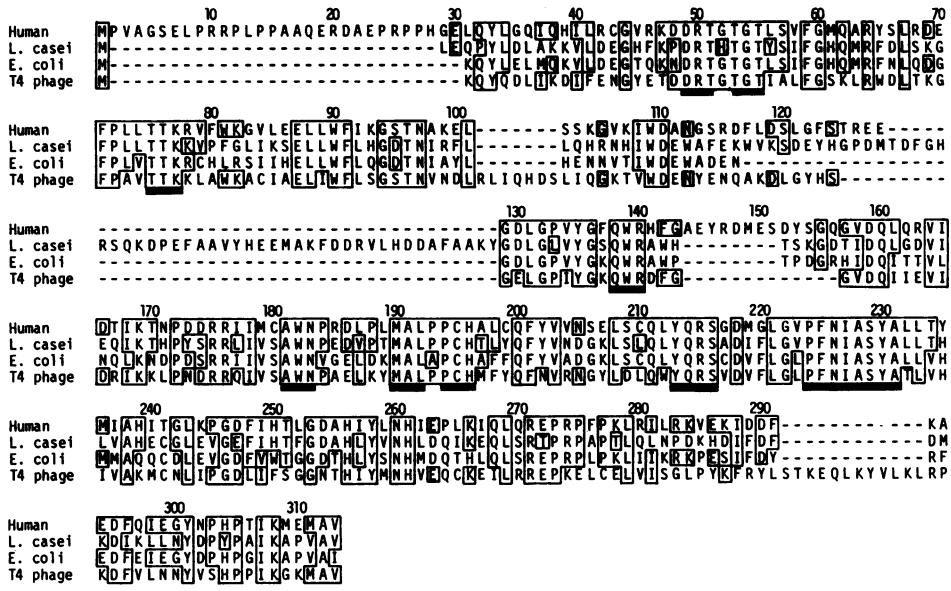

Fig. 3. Comparison of the human and prokaryote thymidylate synthase sequences. The amino acid sequences of the synthase of man (this study, and ref. 19), E. coli (21) and T4 phage (22) were deduced from the corresponding DNA sequences. The amino acid sequence of the L. casei enzyme (20) was determined by amino acid sequence analysis. Amino acids are designated by standard one-letter abbreviations. The prokaryote synthase sequences were aligned with the human synthase sequence to give the best match. The amino acid sequence is numbered for the human enzyme. The amino acid residues that are identical in the human and prokaryote synthase sequences are boxed. Completely conserved sequences in the four species are indicated by solid bars below the four sequences.

frame was completely consistent with that of the first 24 residues which we have determined from the NH_2-terminus of the human synthase purified from mouse transformant cells overproducing human thymidyalte synthase (19). The deduced amino acid sequence was found to have sequence homologous with those of the prokaryote thymidylate synthase in many regions as shown below.

Predicted Amino Acid Sequence

The predicted amino acid sequence of the human synthase was compared with those of the three prokaryotic enzymes, i, e., of Lactobacillus casei (20), E. coli (21) and bacteriophage T4 (22). As seen in Fig. 3, the amino acid sequence has been highly conserved during evolution from bacteria to man. Surprisingly, the homology between the human and bacterial amino acid sequences is comparable to that between the prokaryotes (53% between man and

E. coli and 60% between E. coli and L. casei). In addition, the T4 phage synthase shows about 48% homology with both the human and E. coli enzymes. In contrast, the amino acid sequence homology of the dihydrofolate reductase of man (23) and E. coli (24) is only 25%. Comparison of the amino acid sequences of human and prokaryote thymidylate synthases shows that the human enzyme has about 30 extra amino acid residues in the NH_2-terminal region (Fig. 3). The region thereafter is rather conserved, inclucing some sequences that are exactly the same in all four species; i.e. seven tripeptides, one tetrapeptide and one octapeptide, as specified in Fig. 3. One of these sequences, Thr-Thr-Lys (amino acid positions 75-77), is included in the region identified as folylpolyglutamate binding sites in L. casei thymidylate synthase (25). Another conserved tripeptide sequence, Pro-Cys-His (amino acid positions 194-196), contains Cys-195 which is known to bind 5-fluoro-2'-deoxyuridylate (21,22). The other conserved sequences must also be essential for thymidylate synthase activity.

Untranslated Region

The 5'-untranslated region of human thymidylate synthase mRNA has unique structural features (Fig. 4): First, the G+C content of this region is very high (80%), compared with those in the same region of known mRNA sequences (26). Second, it contains triple tandemly repeated elements immediately upstream of the ATG initiator codon. Two of those elements have exactly the same 28-base sequence and the third element has the same sequence, but with a six-base insertion and a one-base substitution (Fig. 4). This structure is not present in the 5'-untranslated region of any other eukaryotic mRNAs, including the housekeeping and non-housekeeping mRNAs listed in ref. 26, or prokaryotic mRNAs (21,24,27,28) including Escherichia coli thymidylate synthase mRNA. Third, each element of the tandem repeats contains the sequence CGCCGCG (Fig. 4). The three repeats of this sequence in the 5'-untranslated region are complementary with each other. Consequently, the nucleotide sequence in the 5'-untranslated reigon of the synthase cDNA can form three interconvertible secondary structures, each of which contains a stem-loop structure formed by the association of the two CGCCGCG sequences as shown schematically in Fig. 4. This interconvertible structure is of interest in relation to possible translational regulation of thymidylate synthase gene expression by its interaction with some cellular factors.

It remains to be investigated whether this unique structural feature has functional significance. The tandemly repeated structure may not be an artifact formed during cDNA cloning, since other independently isolated cDNA

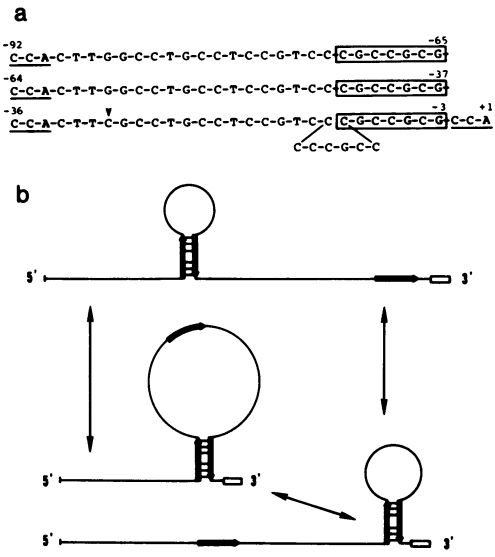

Fig. 4. Unique structural feature of the 5'-untranslated region of human thymidylate synthase cDNA. (a) Sequence comparison of triple tandemly repeated elements. The numbering system of nucleotides is the same as in Fig. 2. The arrowhead indicates the position of base substitution. Inverted repetitious sequences are boxed. The CCA sequences present on both sides of each repeated element are underlined. Because of the presence of this sequence, four kinds of base sequences are possible as a unit sequence of triple tandem repeats. (b) Three possible interconvertible stem-loop structures. The inverted repeat sequence CGCCGCG is indicated by thick arrows. An ATG initiator codon is indicated by an open box.

clones, pcHTS-2 and pcHTS-3 (10), also have exactly the same nucleotide sequence as that of pcHTS-1 in their 5'-untranslated regions, except that pcHTS-3 lacks 10 base pairs at the 5'-end of pcHTS-1 (unpublished data). It is also noted that the plasmid pcHTS-1 is a functional cDNA clone, as mentioned earlier, and the length of the 5'-untranslated region is not unusual (20) and that hypoxanthine-guanine phosphoribosyltransferase cDNA (29) and thymidine kinase cDNA (30) isolated from the same human cDNA library as we used do not have this unique structure.

Two polyadenylation signal (AATAAA) sequences were found in the 3'-untranslated region (nucleotide positions 1126 and 1414), as shown in Fig. 2. We previously observed that a primary transformant 11-1 (FS*thy*-11/*thy*H^+-1) and a secondary transformant D3 (FS*thy*-11/*thy*H^+-D3) produced active human thymidylate synthase mRNA with approx. 1,600 nucleotides, while a secondary

transformant C0 (FS*thy*-11/*thy*H$^+$-C0) produced the mRNA with approx. 1,400 nucleotides (9). The larger mRNA is very similar in size to the cDNA insert in pcHTS-1, while the smaller mRNA is very similar in size to mRNA that might be formed using the polyadenylation signal nucleotide position 1126. In this connection, it is interesting that the dihydrofolate reductase gene of mouse (31), Chinese hamster (32) and human (33) cells produces multiple polyadenylated mRNAs of distinct sizes , presumably formed using different polyadenylation signals in the 3'-untranslated region.

ACKNOWLEDGEMENTS

We thank Dr. Hiroto Okayama, National Institutes of Health and Dr. Paul Berg, Stanford University, for providing the human cDNA library. We also thank Dr. Hiroyuki Sugisaki, Kyoto University, for M13 cloning vector phages and host cells used in the earlier part of this study. This work was supported in part by Grants-in-Aid for Cancer Research and for Special Project Research from the Ministry of Education, Science and Culture of Japan, and from the Society for Promotion of Cancer Research of Japan.

REFERENCES

1. Navalgund, L.G., Rossana, C., Muench, A.J. and Johnson, L.F. (1980) J. Biol. Chem. 255, 7386-7390.
2. Rode, W., Scanlon, K.J., Moronson, B.A. and Bertino, J.R. (1980) J. Biol. Chem. 255, 1305-1311.
3. Ayusawa, D., Shimizu, K., Koyama, H., Takeishi, K. and Seno, T. (1983) J. Biol. Chem. 258, 12448-12454.
4. Hori, T., Ayusawa, D. and Seno, T. (1984) Proc. Jpn. Cancer Assoc. (The 43rd Annual meeting) P. 179.
5. Yunis, J.J. (1983) Science 221, 227-240.
6. LeBeau, M.M. and Rowley, J.D. (1984) Nature 308, 607-608.
7. Hecht, F. and Sutherland, G.R. (1984) Cancer Genet. Cytogenet. 12, 179-181.
8. Ayusawa, D., Shimizu, K., Koyama, H., Takeishi, K. and Seno, T. (1983) J. Biol. Chem. 258, 48-53.
9. Takeishi, K., Ayusawa, D., Kaneda, S., Shimizu, K. and Seno, T. (1984) J. Biochem. (Tokyo). 95, 1477-1483.
10. Ayusawa, D., Takeishi, K., Kaneda, S., Shimizu, K., Koyama, H. and Seno, T. (1984) J. Biol. Chem. 259, 14361-14364.
11. Okayama, H., and Berg, P. (1983) Molec. Cell Biol. 3, 280-289.
12. Kupersztoch-Portnoy, Y.M., Lovett, M.A. and Helinski, D.R. (1974) Biochemistry 13, 5484-5490.
13. Vogelstein, B. and Gillespie, D. (1979) Proc. Natl. Acad. Sci. USA 76, 615-619.
14. Messing, J. (1983) in Methods Enzymol. 101, 20-78.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.

17. Queen, C. and Korn, C.J. (1980) Methods Enzymol. 65, 595-609.
18. Lockshin, A., Moran, R.G. and Danenberg, P.V. (1979) Proc. Natl. Acad. Sci. USA 76, 750-754.
19. Shimizu, K., Ayusawa, D., Takeishi, K. and Seno, T. J. Biochem.(Tokyo) (in press)
20. Maley, G.F., Bellisario, R.L., Guarino, D.U. and Maley, F. (1979) J. Biol. Chem. 254, 1301-1304.
21. Belfort, M., Maley, G., Pedersen-Lane, J. and Maley, F. (1983) Proc. Natl. Sci. USA 80, 4914-4918.
22. Chu, F.K., Maley, G.F., Maley, F. and Belfort, M. (1984) Proc. Natl. Acad. Sci. USA 81, 3049-3053.
23. Masters, J.N. and Attardi, G. (1983) Gene 21, 59-63.
24. Smith, D.R., and Calvo, J.M. (1980) Nucleic Acids. Res. 8, 2255-2274.
25. Maley, G.F., Maley, F. and Baugh, C.M. (1982) Arch. Biochem. Biophys. 216, 551-558.
26. Kozak, M. (1984) Nucleic Acids Res. 12, 857-872.
27. Richardson, K.K., Fostel, J. and Skopek, T.R. (1983) Nucleic Acids Res. 11, 8809-8823.
28. Nakamura, K., Pirtle, R.M., Pirtle, I.L., Takeishi, K. and Inouye. M. (1980) J. Biol. Chem. 255, 210-216.
29. Jolly, D.J., Okayama, H., Berg, P., Esty, A.C., Filpula, D., Bohlen, P., Johnson, G.G., Shively, J.E., Hunkapillar, T. and Friedmann, T. (1983) Proc. Natl. Acad. Sci. USA 80, 477-481.
30. Bradshaw, H.D. and Deininger, P.L. (1984) Molec. Cell. Biol. 4, 2316-2320.
31. Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schimke, R.T. (1980) Cell, 22, 361-370.
32. Melera, P.W., Davide, J.P., Hession, C.A. and Scotto, K.W. (1984) Mol. Cell. Biol. 4, 38-48.
33. Morandi, C., Masters, J.N., Mottes, M. and Attardi, G. (1982) J. Mol. Biol. 156, 583-607.